Government of Canada | Gouvernement du Canada
Canadian Forestry Service | Service canadien des forêts

FIDSINFOBASE: The Forest Insect and Disease Survey Information System

J.M. Power

Information Report PI-X-65
Petawawa National Forestry Institute

87-00080/

PETAWAWA NATIONAL FORESTRY INSTITUTE

In common with the rest of the Canadian Forestry Service, the Petawawa National Forestry Institute has as its objective the promotion of better management and wiser use of Canada's forest resource to the economic and social benefit of all Canadians. Objectives of program activities carried out at the Institute support this goal through discovery, development, demonstration, implementation, and transfer of innovations. Because it is a national institute, particular emphasis is placed on problems that transcend regional boundaries or that require special expertise and equipment that cannot be duplicated in CFS regional establishments. Such research is often performed in close cooperation with staff of the regional centres or provincial forest services.

Research initiatives and technical services at the Institute encompass six major activities:

FOREST GENETICS AND BIOTECHNOLOGY — Integrates projects in tree genetics, soil microbiology, micropropagation, molecular genetics, and seed research. It also includes the client services and seed bank operations of the National Tree Seed Centre, a long-standing program with extensive international affiliations.

FOREST MANAGEMENT SYSTEMS — This program integrates projects in fire, remote sensing, meteorology, modelling, growth and yield, and FIDS to provide research and development for the formulation and demonstration of forest management systems.

NATIONAL FOREST RESOURCE STATISTICS — Widely known as FORSTATS, this program coordinates the acquisition, standardization, and publication of national statistics on the forests of Canada, and responds to client requests.

FOREST DATA SYSTEMS — Provides the expertise and computer-based systems necessary to manage data acquired through FORSTATS and the research projects at PNFI. Additionally, the program provides informatics advice and service at the CFS corporate level.

COMMUNICATIONS — Integrates activities of the library, public awareness, information, and editing and publications projects. The Institute is visited by more than 20 000 people every year. There is a Visitor Centre for the public, self-guided tours, and an extensive education project. The national repository of all scientific and technical publications of the CFS is at PNFI.

THE RESEARCH FOREST — Besides natural stands manipulated in a variety of ways for silvicultural research, the 98 km^2 Petawawa Forest contains extensive areas of plantations dating back six decades. Research plantations are a source of growth and yield data derived from cultural experiments, and they are becoming valuable for pedigreed genetic materials for micropropagation and molecular genetics studies. The forest also offers opportunities for short- and long-term testing of forest management strategies.

FIDSINFOBASE - The Forest Insect and Disease Survey Information System

Information Report PI-X-65

J.M. Power

Petawawa National Forestry Institute
Canadian Forestry Service

1986

©Minister of Supply and Services Canada 1986
Catalogue No. Fo46-11/65-1986E
ISSN 0706-1854
ISBN 0-662-14860-6

Additional copies of this publication can
be obtained from

Technical Information and Distribution Centre
Petawawa National Forestry Institute
Chalk River, Ont.
K0J 1J0

Telephone: 613 589-2880

Cette publication est aussi disponible en français
sous le titre **FIDSINFOBASE - Système d'information
sur le relevé des insectes et des maladies des arbres.**

Contents

v	Abstract/Résumé
1	Introduction
1	The forest insect and disease sampling form
3	The FIDS sampling form data bank
4	Requirements for a new system
5	The concept of relational databases
6	Hierarchical data model
6	Network data model
6	Relational data model
8	Relational database management systems
9	The development of the sampling form component of FIDSINFOBASE
12	The operation of sampling form component of FIDSINFOBASE
12	System features
16	Processing and distributed access
17	System implementation and management
17	Sampling form redesign
18	An integrated information system
20	Conclusion
21	Acknowledgments
21	Literature cited
	Appendices
23	1. FIDS survey form data fields
26	2. FIDSINFOBASE relations

ABSTRACT

A system was developed by the Forest Insect and Disease Survey (FIDS) Technology Development Project at the Petawawa National Forestry Institute (PNFI) to process FIDS sampling form data and, eventually, to process the many other types of data available to FIDS. The system, FIDSINFOBASE, was designed and programmed using the INGRES relational database management system. Various types of forms data entry, modification, analysis, querying, and reporting are available to each of the six regional FIDS units across Canada, to FIDS headquarters, and FIDS-PNFI operations. There is access to a Digital Equipment Corp. VAX computer, on which the system resides, from remote terminals connected through the Canadian Forestry Service regional computer communications networks. Further access is planned via remote microcomputers. The core of this national survey data, both historical and current, is now available in a readily accessible and processable form. The formation of this database is seen as the cornerstone of further developments in the adding of supporting data from other sources, integration with statistical analysis, geographic information and expert systems, and modeling and forecasting. These developments will assist FIDS professional and technical staff in making judicious forest pest management decisions.

Keywords: forest pest management, information systems, databases, Forest Insect and Disease Survey.

RÉSUMÉ

Dans le cadre du projet de développement technologique du Relevé des insectes et des maladies des arbres (RIMA), l'Institut forestier national de Petawawa (IFNP) a mis au point un système permettant de traiter les données des formulaires d'échantillonnage et, éventuellement, beaucoup d'autres types de données auxquelles le RIMA pourra avoir accès. Ce système, appelé FIDSINFOBASE, a été conçu et programmé à l'aide du système de gestion de bases de données relationnelles INGRES. Divers types de saisie, de modification, d'analyse, de consultation et de présentation des données des formulaires peuvent être utilisés par chacune des six sous-sections régionales du RIMA au pays, par l'administration centrale du RIMA ainsi que par le service RIMA de l'IFNP. Grâce à des terminaux régionaux, il est possible d'avoir accès au système, qui réside dans un ordinateur VAX de la Digital Equipment Corp., en passant par les réseaux régionaux de télécommunications du Service canadien des forêts. Les plus importantes données actuelles et historiques sur le relevé national sont maintenant disponibles sous une forme facilement accessible et traitable. La création de cette base de données est considérée comme le point de départ de nouveaux progrès, notamment l'addition de données utiles provenant d'autres sources, l'intégration à l'analyse statistique, à l'information géographique et aux systèmes experts, ainsi que la modélisation et la prévision. Ces améliorations aideront le personnel technique et professionel du RIMA à prendre des décisions judicieuses en ce qui concerne la répression des ravageurs forestiers.

Mots-clés: répression des ravageurs forestiers, systèmes d'information, bases de données, Relevé des insectes et des maladies des arbres.

FIDSINFOBASE - The Forest Insect and Disease Survey Information System

INTRODUCTION

The Forest Insect and Disease Survey (FIDS) of the Canadian Forestry Service (CFS) is a national agency involved in the surveillance and analysis of forest pest conditions, and in the dissemination of the results to the forest management community. A major task of this agency is the gathering and processing of data, and transfer of information. Mechanization of components of this "system" is a function which FIDS has undertaken in the past, and continues to do.

A new effort to computerize the FIDS sampling form databank was undertaken by the FIDS Technology Development Project, PI-20, at the Petawawa National Forestry Institute (PNFI). This databank forms the main source of historical data on forest pest conditions in Canada. However, it only represents a fraction of the data that exists within FIDS in either digital, paper, or cerebral form. The development of a new system which handles sampling form data takes into account the possibility that the remaining vast quantity of data will also be brought on stream in the future. This report concentrates on the component which now handles the sampling form databank, known as FIDS-INFOBASE. It is the initial component of a system that will evolve into a comprehensive forest pest management information system.

The FIDS sampling form

A major function of FIDS is to conduct ground surveys to detect and confirm the existence and extent of forest pest infestations. These surveys involve observing the condition of the forest and taking samples of plant material and pests for analysis. For the past four decades, FIDS and its predecessor, the Forest Biology Division of the Department of Agriculture, have recorded field observations and subsequent laboratory analysis results on forms. These forms are known as FIDS "enclosure slips" or sampling forms. Before 1966, separate forms were used to record field observations associated with insects or diseases. The data consisted of a basic description of the stand type, the sampling technique, and the location by township or other administrative unit.

A major redesign took place during 1966, which allowed for more detailed descriptions of the field data plus attributes associated with the laboratory taxonomic process (Figure 1). One form handled both insect and disease survey results. The form was also suitable for keypunching because all data were coded, except for the remarks. To assure national standardization across FIDS units, and to assist in coding, the "Data Recording and Retrieval Manual" (Brown and Davidson 1968) was provided as a guide.

Aside from the introduction of metric units into the sampling form in 1984, it has not changed in appearance since 1966. However, due to increased demands to accommodate more types of descriptions and variations of the original survey, attributes on certain fields have been modified on a regional basis. The form has two sides: the front is used to record field

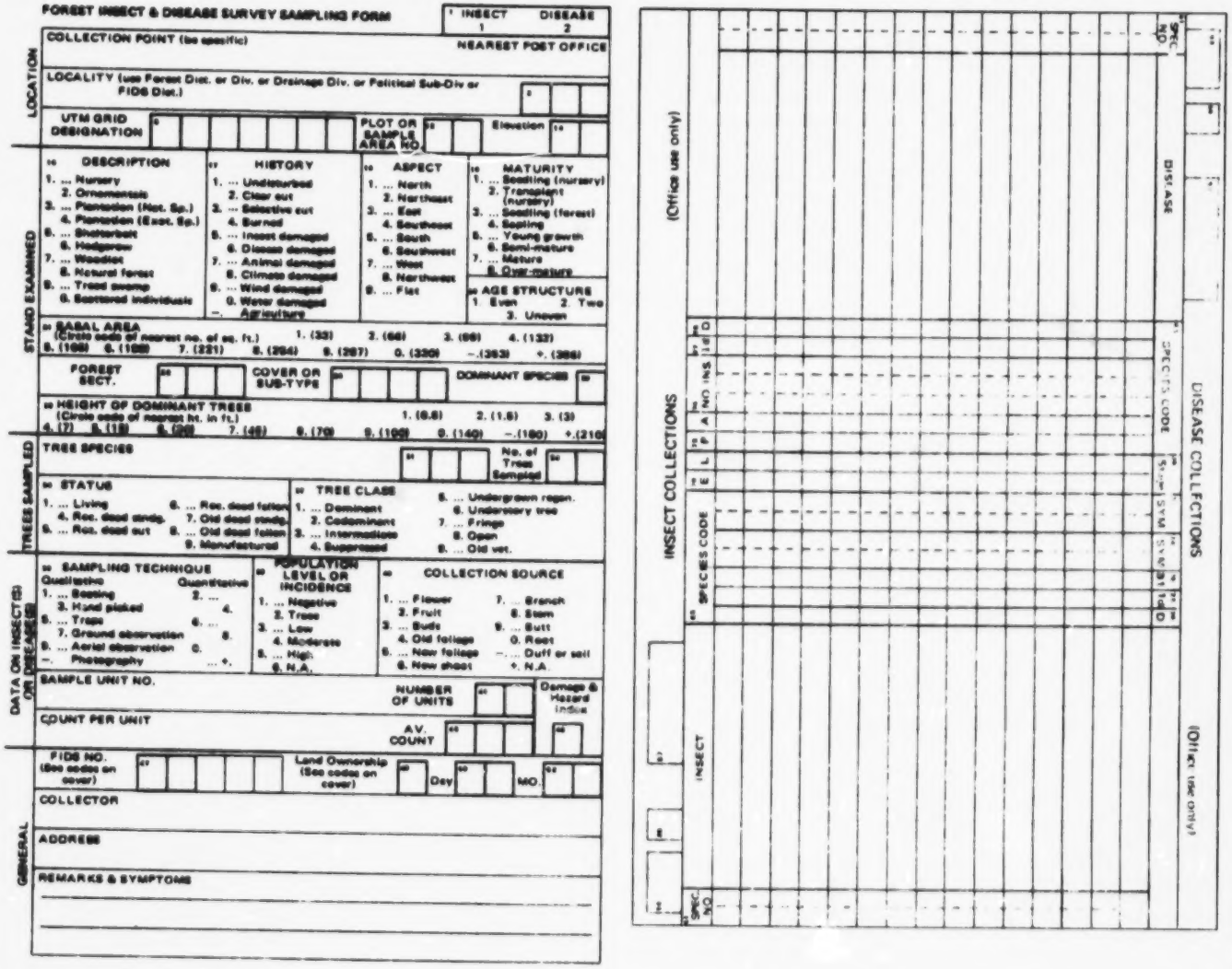


Figure 1. Forest Insect and Disease Survey - "Enclosure Slip" on Sampling Form.

observations and the rear is for recording laboratory observations of pest specimens. The data that are recorded are: location and date within which the sampling or observations took place; surrounding physiographic features; a description of the forest stand, host trees, and pest action; sampling procedure; taxonomic identification; and, characteristics of the specimen. Appendix 1 has a description of each data field.

A sampling form is filled out for the survey for each host tree species for each survey event, for either the insects or diseases found on that host. Any number of insect species or diseases can be represented by each form. Insect and disease collections are never mixed on the same form.

The field data portion of the form is completed during the survey event. A copy is retained by the FIDS ranger for his records. The laboratory data portion is complete following positive identification. The time required for identification varies, because some samples may require a lengthy rearing or culturing period. The identification results are retained by the laboratory and the FIDS ranger.

The FIDS sampling form databank

The 1966 redesign of the sampling form permitted additional data to be recorded, such as that on the forest stand under examination. A system was developed to accept keypunched data from these forms, store them in a databank, and allow various reports to be generated. The objectives of this databank were:

A. to maintain historical records of forest pest specimen surveys,

B. to provide regional and national data to the regions, and

C. to provide biological data from research purposes.

The databank resided on an IBM mainframe computer at DataCrown Ltd. in Toronto. The system was developed by the Computing and Applied Systems Directorate of Environment Canada (CASD) using IBM data management software (Mark IV). It was maintained by them on behalf of FIDS until 1985. At that point the system was temporarily taken over by the Systems and Consulting Directorate (SCD) of Agriculture Canada, which gave it to PNFI in 1986.

The survey forms were shipped annually to CASD for keypunching and entry into the Mark IV system through batch jobs. Edit and accession listings were sent to the various FIDS units for data correction. Annual reports and specially-requested historical reports were then produced.

A major effort was undertaken to recode and input all of the pre-1966 sampling forms into the databank. Although many discrepancies existed in data due to conversion, a valuable source of digital information on forest pests dating back to the 1930s was obtained. As of 1984, the data bank contained 1.2 million records, representing 600 000 survey forms. Currently, new forms are added at a rate of 10 000 per year.

Associated with the sampling form records is a separate file which lists every forest insect and disease species with which FIDS has dealt. Each

record in the file contains the Latin binomial for each species, a species code, and a pointer code that links the record to other synonymous species. Modifications are made as new species are found, or as taxonomists determine synonymity between two species present in the file. New insect and disease catalogues are regularly printed and distributed to FIDS units for reference, in particular for coding species on the survey forms.

The databank is the only major computer file of information from which FIDS scientists can conduct analyses on forest pests, their presence, their action on the forest, and their trends. It is a vital source to these scientists for pest research, survey planning, and for information transfer to forest managers on pest situations.

Requirements for a new system

FIDS saw a need to improve access to the system for quicker entry of new forms and for receiving reports, and to facilitate retrieval of more complex information. The new FIDS Technology Development Project at PNFI was given the task of developing a new system to handle sampling form data.

The new project confronted several problems with the old databank:

A. It was based on old technology, requiring batch access with extensive use of magnetic tapes.

B. It was not directly controlled by FIDS, which increased reaction time for data processing, and reduced FIDS's ability to have direct input on data management and keeping up with new developments.

C. Summarization and analysis capabilities were limited or nonexistent. New requirements for using mapping or statistical analysis software with the data were not feasible.

D. Turnaround time for data requests ranged from two weeks to six months. This resulted in limited use by the regions and discouraged many potential retrievals.

E. Poor access meant that many errors were not corrected. Any exercise by regional FIDS staff to improve the integrity of data was not feasible.

F. The system was not able to handle an evolving and growing survey with respect to the changing data elements in fields, and the incorporation of new types of data.

G. Documentation was poor, because the system was old and had changed hands many times.

A study was done on the feasibility of improving the system and bringing it under FIDS control and management. The assistance of a contractor, DMR and Associates, was retained to bring data modelling and data management expertise to the study. The study resulted in a proposal (DMR and Assoc. 1985) for a system with features that would:

i. Provide user control of the data because regional FIDS units collect the data and are major users.

ii. Enable user data entry for on-site error correction and better access to current data.

iii. Facilitate user-friendly access to the data, which meant flexibility to accommodate special and standard requests.

iv. Allow for expansion of data as the survey evolves, and for integration with other pest data for an eventual comprehensive forest pest management information system.

v. Provide for regional and national levels of data.

The system was developed with an emphasis on establishing a resource information base for FIDS. It was structured around a relational database using tools to build up applications, and satisfy queries and reporting requirements in a short time frame. A prototyping approach was used so that the system infrastructure would be in place quickly, followed by tuning and expansion at later stages as patterns of user operation become well-defined. The extant prototype gives users the opportunity to get hands-on experience.

The relational database management system (RDBMS), INGRES, was confirmed as a suitable development tool for the FIDS sampling form database. INGRES, which is the creation of Resource Technology Inc., utilizes a fourth generation language for application development, querying, and reporting. INGRES was already in use by Agriculture Canada and was recommended by the contractor as being one of the top five relational database management systems able to operate on the PNFI and CFS computer architecture. INGRES was already in use as a pest management database in Michigan (Russell et al. 1984), on a Digital Equipment Corp. (DEC) VAX 11/730. The DEC VAX 11/780 minicomputer at PNFI was used for system development and initial operation. Computer hardware was an important factor, as a RDBMS requires large amounts of memory, mass storage, and CPU cycles in place of time-consuming systems analysis and programming. The investment in relational database technology for this task was viewed to potentially yield benefits in computing as part of the FIDS endeavour towards automation and towards the development of forest pest management systems in general.

The concept of relational databases

The methods used in digital storage and access of data become more critical as data files, the complexity of retrievals, and performance requirements increase. Database management systems are tools which assist in the organization and manipulation of data through careful design of storage and access procedures. This design is known as the "data model". The three basic models used in database management systems are hierarchical, network, and relational, and are described by Hasling (1985).

Hierarchical data model

This model, the original used in computing, is characterized by a hierarchical ordering of data files. In an example using FIDS sampling forms, a file may exist which lists all insects found in each plot (Figure 2). A different way of storing this information would be by listing all plots where each insect was found. Now, say that we want to know if insect x occurs in plot 1. In the first case, it would be easy to find all of the insects in plot 1. But in the second case, retrieval would be a tedious operation, because every record for each insect would have to be searched to see if plot 1 occurs. In this model, one must decide on the type of retrievals that will be made from the database before records are designed. Otherwise, the result will be a loss of efficiency. Data redundancy is another problem of retrieval inflexibility. For example, all of the attributes for plot 1, such as plot location, would have to be duplicated for each insect found in the plot because a separate record would exist for each insect. Not only does this increase storage requirements, but causes the database to become inconsistent and outdated. For example, changes in attributes for a particular insect may erroneously be not reflected in every occurrence of that insect. In this model, plots and insects have a many-to-many relationship which, as we see, is not handled well. Hierarchical models work best with one-to-many relationships.

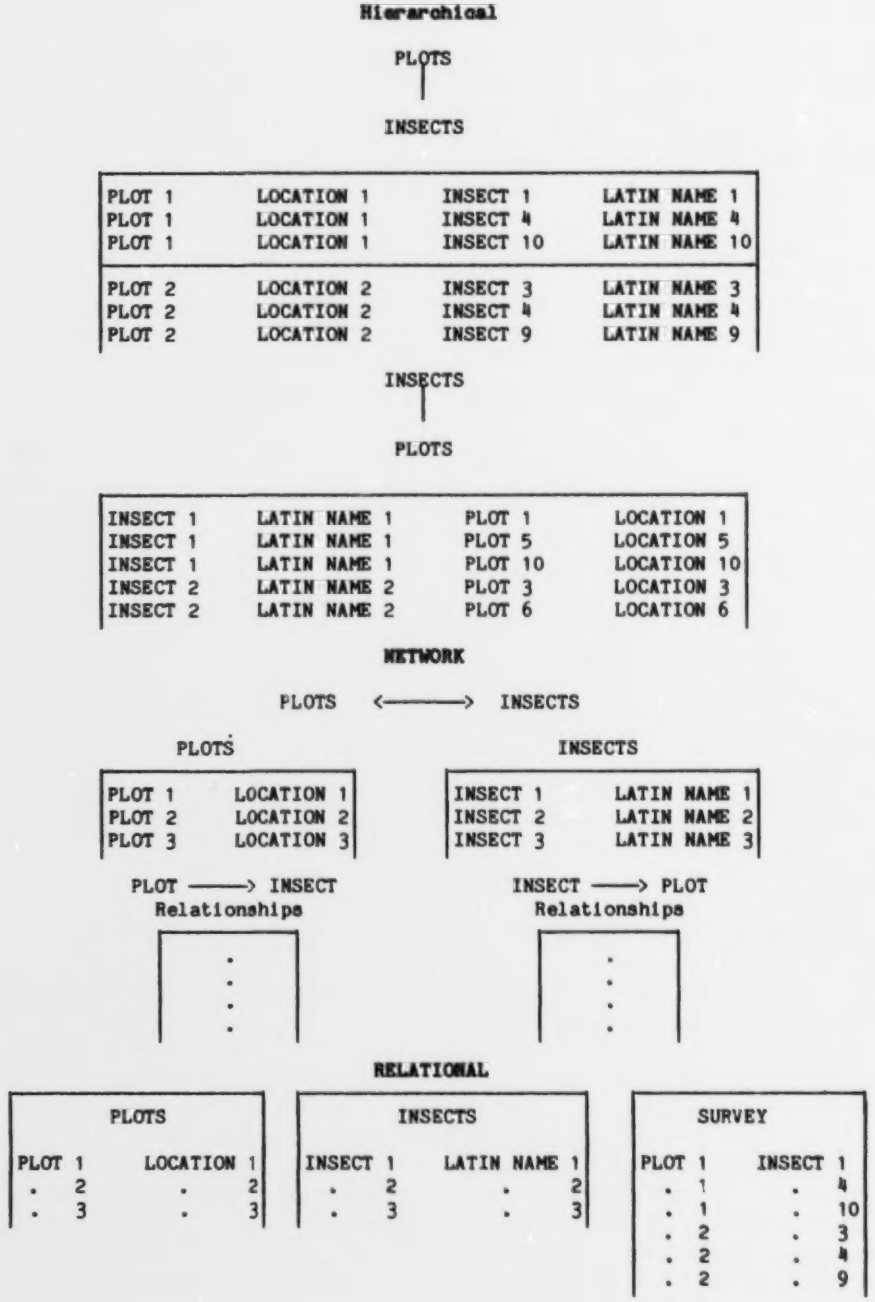
Network data model

In the network data model, the problems of data redundancy and many-to-many relationships are eliminated. Each entity is represented by an unique file (Figure 2). Files are networked by identifying links which connect them. In our example, plots and insects would be listed in separate files. In addition, tables of links would exist which identify the plot-insect relationships and the insect-plot relationships. Therefore, a query to find all insects in plot X would use the plot-insect relationship table to locate the link which would help navigate from the plot file to the insect file so as to retrieve the desired data efficiently. A large number of computing procedures are required in each query due to the complexity of the operation. This model works well where the database is stable and retrievals are well established; however, to change data or modify the network links would be a major task.

Relational data model

In this model all information is stored in the form of two-dimensional tables or "relations". As shown in Figure 2, tables contain key fields, such as plot number and insect code, and non-key fields, such as location and the Latin binomial of the insect species. The data in these tables are normalized to create independence; for example, the plot table would not contain attributes of a particular insect found in the plot. These would be contained in an insect table. The survey table would list the insects and plots associated with each other. The order of rows are irrelevant. Normalization is a logical database design method used in relational databases. In what are known as "second and third normal forms", every field in a record is either part of the key field or provides a single-valued fact about exactly the whole key and nothing else (Kent 1983).

Figure 2. Data models.


Hierarchical
PLOTS
INSECTS
PLOT 1 LOCATION 1 INSECT 1 LATIN NAME 1
PLOT 1 LOCATION 1 INSECT 4 LATIN NAME 4
PLOT 1 LOCATION 1 INSECT 10 LATIN NAME 10
PLOT 2 LOCATION 2 INSECT 3 LATIN NAME 3
PLOT 2 LOCATION 2 INSECT 4 LATIN NAME 4
PLOT 2 LOCATION 2 INSECT 9 LATIN NAME 9
INSECTS
PLOTS
INSECT 1 LATIN NAME 1 PLOT 1 LOCATION 1
INSECT 1 LATIN NAME 1 PLOT 5 LOCATION 5
INSECT 1 LATIN NAME 1 PLOT 10 LOCATION 10
INSECT 2 LATIN NAME 2 PLOT 3 LOCATION 3
INSECT 2 LATIN NAME 2 PLOT 6 LOCATION 6
NETWORK
PLOTS <——> INSECTS
PLOTS
PLOT 1 LOCATION 1
PLOT 2 LOCATION 2
PLOT 3 LOCATION 3
INSECTS
INSECT 1 LATIN NAME 1
INSECT 2 LATIN NAME 2
INSECT 3 LATIN NAME 3
PLOT ——> INSECT Relationships
INSECT ——> PLOT Relationships
RELATIONAL
PLOTS
PLOT 1 LOCATION 1
. 2 . 2
. 3 . 3
INSECTS
INSECT 1 LATIN NAME 1
. 2 . 2
. 3 . 3
SURVEY
PLOT 1 INSECT 1
. 1 . 4
. 1 . 10
. 2 . 3
. 2 . 4
. 2 . 9

The basic rules (Sandberg 1981) in this model are:

- each table contains only one record type
- each record has a fixed number of fields
- fields are distinct so that repeating groups are not allowed
- there is no predetermined sequence in records
- fields take their values from a domain of possible field values.

Codd (1982) describes the relational model in three parts: structural, manipulative, and integrity. The structural part consists of domains, relations (tables) of assorted degrees (columns in a table), attributes, tuples (rows), candidate keys, and primary keys. The order of columns and rows in tables are arbitrary and irrelevant.

The manipulative part consists of algebraic operators such as select, project, and join, which transform relations into new relations. The select operator selects certain rows in a table to build a new table. The project operator selects certain columns in a row to build a new table. This table may have fewer rows owing to the possibility of having duplicate row values after the selection. Duplicate rows are weeded out. The join operator merges two tables on the basis of values from one column in each table. The two columns must contain field values that are comparable, in that they are from the same domain. Join operations are tedious, but can be made more efficient through query-optimization techniques (Sandberg 1981).

The integrity part of the model consists of setting rules which define the state or changes in state of the database, such as those which govern attribute definitions and keys.

Relational database management systems

There are many database management systems available which are based on one of the above models or combinations of all three. The rules which identify a DBMS as being "relational" are that they contain (Codd 1982):

- tables without user-visible navigation links between them
- and a data sublanguage with a relational processing capability.

Many of the so-called relational database management systems on the market may actually be nonrelational or tabular; to qualify they have to meet the above rules. Relational database management systems are usually integrated with a high-level "fourth generation" language that allows system development and user-generated retrievals to be nonprocedural. This reduces the human effort on the system. The term "fourth generation" is applied to the newer programming languages which can be quickly learned and easily operated by non-technical users. System development can be achieved without the traditional elaborate analytical process of "getting it right the first time" before any development starts. A prototype can be placed in the hands of the user soon after the conception of the system. Users who are inexperienced in computing can design their own queries, reports, and data entry and manipulation processes.

Codd (1982) suggests that a full-scale DBMS should have the following capabilities:

- data storage, retrieval, and update
- an user-accessible catalogue for data description
- transaction support to ensure that all or none of a sequence of changes are reflected in the pertinent databases
- recovery services in case of failure
- services that control transactions that occur concurrently
- authorization services for security
- integration with support for data communications
- integrity services.

The INGRES RDBMS selected for the development of the sampling form database best approached these capabilities. Written in the C language, it incorporates the data sublanguage QUEL for use interactively or as part of a host language (e.g. FORTRAN) application program (RTI 1985a). INGRES was developed as a full-function multi-user RDBMS. It is currently the most popular RDBMS on VAX computer systems (RTI 1985b).

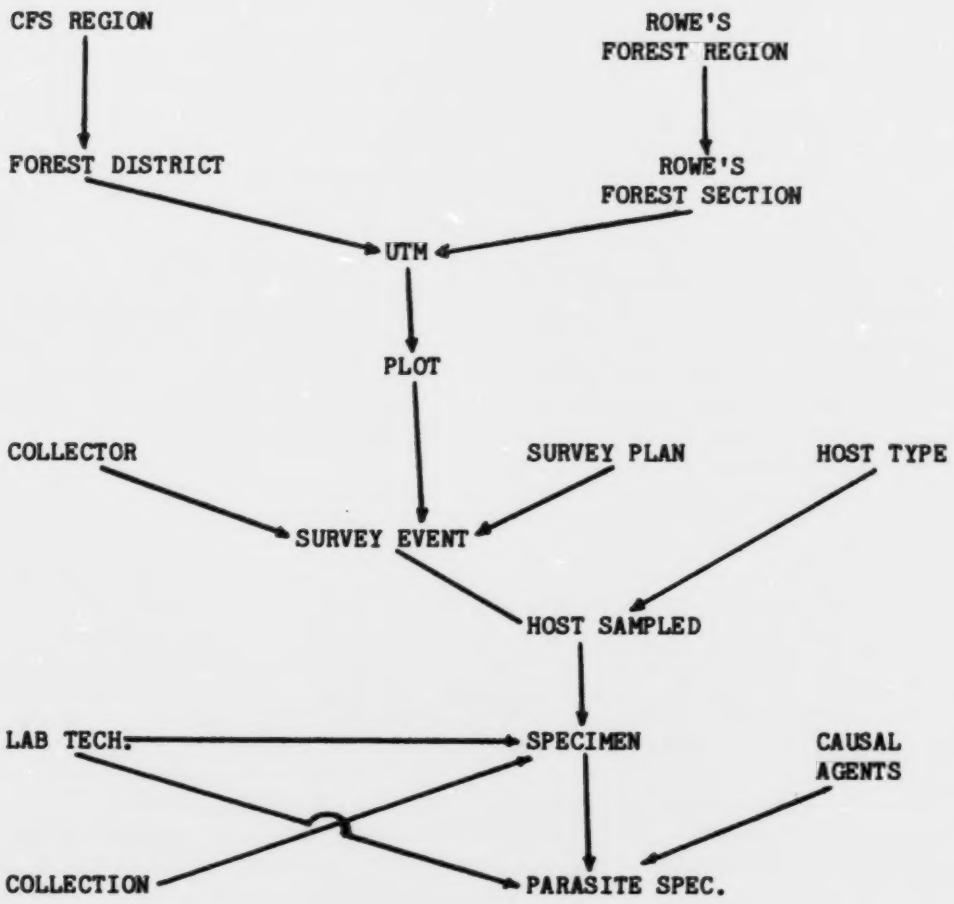
The development of the sampling form component of FIDSINFOBASE

DMR and Associates was contracted to develop a prototype sampling form database. The components of the original sampling form databank and the more current uses of the survey form were analyzed in order to understand various survey data entities and their interrelationships. The data model from which the database was designed was made using this knowledge. The model (Figure 3) contains a number of entities and their interrelationships, most of which are represented by the existing attributes on the survey form. Entities are linked to each other by "one-to-one" or "one-to-many" relations.

A guide through the model will assist in comprehending its structure. The core entity could be the geographic location, which is the Universal Transverse Mercator (UTM) coordinate system adopted by FIDS. UTMs are contained within administrative units such forest districts, which in turn are contained within the CFS regions. On the physiographic side, UTMs are contained within Rowe's forest regions and sections.

Within UTMs are the survey sample areas, which are referred to as "plots". Plots may or may not have a defined area but represent the locality to which the survey measurements apply. In many cases plots only exist for the duration of the survey; in others, remeasurements are intended, so it is important to retain a temporal aspect for each plot in the form of a survey event entity. Survey events are planned in some manner in advance; these plans may not be recorded at present on the form, but are considered an entity. Survey events are conducted by survey rangers or other collectors. During each event, one or more host tree species are sampled for pests. At this point a separate survey form is used for each host tree species for either insect or disease information. The sampling of host trees may result in the collection of pest specimens, or material which may contain pests. These specimens could later be analyzed by laboratory technicians, identified as one of the "causal agents" located in a catalogue, and perhaps stored in a permanent collection. In addition, specimens may contain parasites, which in themselves may form new specimens. Parasites were accommodated because the PFC FIDS unit was already recording this information supplementary to the survey form data.

Figure 3. Data Model of Sampling Form Component of FIDSINFOBASE



The entities and their relations will have associated attribute data which are the data elements found on survey forms. Element descriptions form part of the "data dictionary" in the FIDSINFOBASE database. The data dictionary and associated catalogues and check tables were loaded into INGRES in the form of tables or relations. These relations and their contents are listed in Appendix 2. The catalogues and check tables are static components of the database which show valid data values, data codes, and other features such as the synonymity between certain species in the catalogues. Procedures were then developed using INGRES to allow loading in modifying data, browsing through data files, searching and selecting data through queries, and producing reports.

Various components are available in INGRES for system development:

(i) QUEL - the INGRES querying language.

(ii) QBF - Query-By-Forms, a forms-based querying function for the average user. A form consists of an organized layout, of prompts for information from the user, with adjacent space allocated to display the keyed-in user response. It provides an easy and friendly approach for an INGRES task to be performed.

(iii) RBF - Report-By-Forms, a forms-based report generator for the average user.

(iv) Report Writer - the INGRES report writing language used for highly customized reports that cannot be prepared using RBF.

(v) Vifred - Visual-Forms-Editor, used to edit screens for readability and ease of use.

(vi) ABF - Application-By-Forms, used to develop and test applications.

Certain tools such as ABF can be used to quickly put together an application for operational use. This is ideal for prototyping -- where one would like to see a system perform the basic functions first, before going into the fine details. If certain components of the system require improved performance, such as on turnaround time for an interactive query, further development can be invested by redesigning the function using a programming language such as FORTRAN and combining it with QUEL procedures.

The FIDSINFOBASE prototype was developed using both fourth generation tools, such as ABF, and FORTRAN/QUEL routines. As the system evolves, components can be tuned and new ones added.

The database was divided to handle two types of data: historical - all of the data existing in the old data bank; and, new - any data that is entered interactively into FIDSINFOBASE. The new data would pass through all of the error-checks in INGRES and then be placed into their appropriate relations. Loading of these data can be done interactively or in batch form. The historical data require a significant amount of cleaning before they can be loaded into the database in a fully-relational form. They will be treated as a separate, single relation so that they can be accessed and corrected with INGRES tools by the FIDS regional staff.

The operation of the sampling form component of FIDSINFOBASE

System features

FIDSINFOBASE is entered by an authorized user through the selection of functions on menus which are displayed on a video terminal connected to the system. These functions allow entry of new survey form data, maintenance of survey data already in the database, maintenance of static tables, browsing and querying survey data and static tables, and production of reports. The system permits advanced users to design custom-tailored data entry, browse, query, and reporting, beyond that available on the menu.

As the system evolves, the menus will be improved and expanded. The menu structure as it was in late 1985 will be used to explain the available procedures and their selection.

The main menu procedures are:

A. CODES - Maintenance of translation code tables, which are stable, look-up tables such as insect and disease catalogues, and forest section, cover type, host species, and collector information. A menu is displayed which permits access and modification of any of the code tables. These tables include description fields in both English and French for each code.

B. TABLES - Browse translation code tables. The function is similar to CODES but only allows the tables to be browsed through querying.

C. INSECT CAT - Browse of insect catalogue data. This is a taxonomic list of all forest insect species in Canada encountered and recognized by FIDS. Through querying, individual or groups of catalogue entries that meet the query criteria can be displayed. The insect code with the associated Latin name is shown. Queries may be done by identifying individual codes, optionally accompanied by Boolean operators such as < or >, or with wildcards. For character data, all or part of the name accompanied with wildcard extensions can be used. Two examples are shown in Figure 4. The first is a query for all entries with the character string "ABIETINUS" in the species field. The second example is a query for all entries that have "0000" as the last four figures of the code which, incidentally, will list the taxonomic order and family names. The first few entries of the result are shown (Figure 4).

D. DISEASE CAT - Browse of disease catalogue data. This function is similar to INSECT CAT.

E. SYNONYM - Query taxonomic code synonyms. This function displays all synonymous insect or disease species codes for a selected code. This is essential information for use in conducting queries on the system for a pest species.

F. SURVEY ENTRY - Entry of field survey data. A form is displayed which is used to enter data interactively from the field portion of the survey, and optionally the lab results if these are available at the time of entry.

Figure 4. Examples of Insect Catalogue Queries

Forest Insect and Disease Survey Program
Insect Catalogue

Code	Class	Species
4012501	MINDARUS	ABIETINUS KOCH.
4140308	PINEUS	ABIETINUS UNDER. AND BALSH

Help Query Scroll Exit

Forest Insect and Disease Survey Program
Insect Catalogue

Code	Class	Species
0010000	PSOCIDAE	
0020000	ATROPIDAE	
0030000	LIPOSCELIDAE	
0040000	LACHESCILLIDAE	
0050000	THOMISIDAE	
0210000	TRICHODECTIDAE	
0250000	PHLOEOTHRIPIDAE	
0350000	ENTOMOBRYIDAE	
0360000	PODURIDAE	
0370000	SMINTHURIDAE	
0410000	DIPLOPODA	(MILLIPEDES)
0500000	FORFICULIDAE	
0510000	LABIIDAE	

Help Query Scroll Exit

Each entry is error checked against valid ranges of codes. The system automatically checks the database for a duplicate entry.

G. LAB ENTRY - Entry of laboratory results. This function is used if the field data have been entered on a previous occasion, and the lab results were pending taxonomic analysis. A form is displayed which is used to enter the results of laboratory analysis of pest specimens pertaining to a specific field survey. Each entry is error-checked. The system automatically finds the respective field survey portion for matching.

H. FIELD MAINT - Maintain field survey data. This function allows the user to modify field survey data that is currently in the database.

I. LAB MAINT - Maintain lab results. As in FIELD MAINT, the lab results can be modified with this procedure.

J. SURVEY QRY - Query survey data. Both the field and lab results portions of the survey form are displayed for querying on any of the fields. All forms which met the query criteria are found and are made available for browsing.

K. QBF - Access to Query-By-Forms. Allows the user to develop special queries on the database which may not be offered in the other functions.

L. REPORTS - Selection of the reports subsystem. This function displays another menu of various types of reports available to the user. These reports provide a hardcopy listing of certain data presented in a designed format. The reports are printed out at a printer which can be designated by the user. The types of reports available will increase as the system evolves to further meet the users' needs. Examples of report output shown in Figure 5.

M. RBF - Access to Report-By-Forms. A tool for the user to design his own report generators for the database.

N. QUEL - Access to QUEL. For proficient users to undertake complex database operations which are not available elsewhere on the menu.

O. MAIL - Access to the VAX VMS Mail facility.

P. SHELL - Takes the user out of INGRES and into the VMS command mode.

Each menu screen has a selection of "function" keys which are displayed. These functions invoke a certain action on the procedure selected by the cursor on the menu. These actions vary depending on the menu. The main menu functions are: "GO", which invokes the selected menu item; "Exit", which takes the user out of the menu and up to a higher level in the system, (in this case it exits INGRES): and "Help", which provides the user with instructions on the menu procedures available.

Database integrity is handled in a number of ways in the system to assure that the data is as reliable as possible. Many procedures are powerful in terms of their effect on the database; therefore, it is desirable to allocate access to certain procedures by specified users to avoid unauthorized database

Figure 5. Report examples. - 15 -

```
DATE: 86/03/06                                                                                                    PAGE   1
                                                REPORT -- PIDS-PNPI
                        LISTING OF SURVEY DATA ON PESTS WITHIN A LOCALITY, SORTED BY PEST NAME, TREE SPECIES AND UTM

REGION: 2    LOCALITY: 302   SURVEY TYPE: DISEASES

 U T M      ELEV DES HIS ASP MAT AGE BA ROOT SUB1 SUB2 MGT STA TCL POP OU SP# DISEASE DD MM YY  REG#   HOST
ZI EE NNNN                                                                    SIS2DDS                  # NAME


     PEST SPECIES: ARMILLARIA          MELLEA (PVAHL:FR.) KUMMER                    TAXONOMIC CODE: 3004001

2046 513     00  0  1  8  4  3      009 99 00                    4  0 01  0011001 07 08 80 0343  003 DF
2046 514     00  0  1  7  4  3  4   009 21 00  4  4  2  4  4 01  0911003 06 08 80 0349  003 DF

     PEST SPECIES: BOTRYTIS                                                       TAXONOMIC CODE: 5022000

2050 514     00  4  1  9  5  1  1  009 61 00  7  1  7  4  4 01  0104003 06 08 80 0348  325

     PEST SPECIES: CIBORINIA           WHETZELII (SEAVER) SEAVER                  TAXONOMIC CODE: 2057005

2050 514     00  0  1  9  5  3  3  009 61 00  7  1  7  4  4 01  0706071 06 08 80 0306  311 TA

     PEST SPECIES: COLEOSPORIUM        ASTERUM (DIET.) SYD.                       TAXONOMIC CODE: 4003006

2046 512     00  3  1  9  5  1  2  009 03 00  5  1  2  4  2 01  0507003 04 07 80 0163  060 HF

     PEST SPECIES: CRONARTIUM          RIBICOLA J. C. FISCH.                      TAXONOMIC CODE: 4004007

2046 512     00  3  1  9  5  1  3  009 02 00  5  1  2  4  2 01  0000001 04 07 80 0178  062 EWP
2046 512     00  3  1  9  5  1     009 02 00     1  2  4  2 01  0007001 14 08 80 0570  062 EWP

     PEST SPECIES: HYPOXYLON           MAMMATUM (WAHL.) J.M. MILLER               TAXONOMIC CODE: 2159010

2050 514     00  0  1  9  5  3  3  009 61 00  7  1  7  4  4 01  3013003 06 08 80 0331  311 TA

     PEST SPECIES: MELAMPSORA          MEDUSAE THUEM.                             TAXONOMIC CODE: 4012014

2052 509     00  0  1  9  4  3  1  009 11 61  6  1  4  4  4 01  0106003 07 08 80 0343  311 TA

     PEST SPECIES: MOISTURE LOW        WINTER DRYING                              TAXONOMIC CODE: 7004204

2048 512     00  4  1  9  5  1     009 01 00     1  3  5  4 01  0007001 28 05 80 0556  062 SF

     PEST SPECIES: NECTRIA             COCCINEA VAR. PAGINATA LOHM., WATS.        TAXONOMIC CODE: 2208003
                                       + AVERS

2046 513     00  0  1  1  4  3  2  003 02 93  7  1  3  4  9 01  0000001 27 05 80 0044  240 AR
```

```
DATE:  86/03/14                                                                                                   PAGE    1
                                                REPORT--PIDS-PNPI
    FREQUENCY OF COLLECTIONS IN A SPECIFIED BLOCK OF UTM'S  FOR EACH YEAR FROM 1945 TO 1984, BY PEST NAME

REGION:  2    UTM ZONE:  19    UTM EASTINGS FROM:  69    TO:  71    UTM NORTHINGS FROM:  509    TO:  511

   SURVEY TYPE:  DISEASES    SPECIMEN NUMBERS RANGE FROM:  01    TO:  89

    PEST NAME                                          FREQUENCY OF COLLECTIONS BY YEAR

                                                  84 83 82 81 80 79 78 77 76 75 74 73 72 71 70 69 68 67 66 65
                                                  64 63 62 61 60 59 58 57 56 55 54 53 52 51 50 49 48 47 46 45

AGARICALES                                                                                    1
-------------------------------------------------------------------------------------------------------------------
ALTERNARIA                                                                                       1
TENUIS AUCT.
-------------------------------------------------------------------------------------------------------------------
ARCEUTHOBIUM                                                                                                      1
PUSILLUM PK.                                  1
-------------------------------------------------------------------------------------------------------------------
ARMILLARIA                                       1          1                    1              9
MELLEA (PVAHL:FR.) KUMMER                     1
-------------------------------------------------------------------------------------------------------------------
ASCOCALYX
ABIETIS NAUM.                                                 1          1
-------------------------------------------------------------------------------------------------------------------
ASCOMYCETES                                                                                           1
-------------------------------------------------------------------------------------------------------------------
ATMOS IMPURITIES                                              1                                       1
HERBICIDES
-------------------------------------------------------------------------------------------------------------------
AUREOBASIDIUM                                                    1
PULLULANS (DE BARY) ARNAUD
-------------------------------------------------------------------------------------------------------------------
BACTERIA                                                         1
-------------------------------------------------------------------------------------------------------------------
BASIDIOMYCETES                                                                                        2  1
OOOTHER1OO
-------------------------------------------------------------------------------------------------------------------
```

corruption or destruction. The system manager can restrict access to procedures and to data whose values are within designated parameters for each user. This will prevent unauthorized manipulation. Despite measures to maintain control of the database, there is always the possibility that corrupt data will be entered, or data may be lost due to an user error or system crash. The system contains backtracking features which the system manager can use to restore a database to its correct state.

Follow-on development with the FIDSINFOBASE prototype will include continual improvement and addition of database access after consideration of user reactions. An important addition will be the ability to input spatial coordinates to aid in querying, and to output coordinates for graphic display. For regional operations, a link to a desk-top microcomputer-assisted drafting (CAD) system can be built so that UTM coordinates can be identified through digitizing for a FIDSINFOBASE query, and for the spatial results of a query to be plotted out. Systems under investigation are

- Environmental Systems Research Institute (ESRI) PC ARC/INFO
- Digital Resource Systems Ltd. CAD/GIS
- Aeronca VGS300 Plus

Further use of microcomputers in relation to FIDSINFOBASE are discussed in Section 7.2. In the near future at PNFI, a link to the resident geographic information system (GIS), ESRI ARC/INFO, will enable graphic functions to be performed, as well as more complex spatial processing needs.

The prototype was designed to enable addition of both French and English terms. Text will be available either in bilingual form or with the choice of selecting either French or English.

Processing and distributed access

For development and operation, INGRES requires sufficient computer resources so that its performance, in particular with respect to queries, is satisfactory. Ideally, it should have its own computer to operate with large applications such as FIDSINFOBASE. An investigation on the hardware requirements for FIDSINFOBASE resulted in a recommendation to acquire a dedicated supermicrocomputer, the DEC MicroVAXII (Power and O'Brien 1986). It is expected that an average of six users can access FIDSINFOBASE concurrently on this system. Users can be stationed either locally or connected from remote sites across the country. For 1986, 450 MB of mass storage and up to 9 MB of memory will be required for operations and development.

Usually in the entry, browse, and query modes, the user will be dealing with a batch of data. This would not normally require working with the entire database all of the time. The placement of an user-dedicated computer between the user and the host system would allow working with these database subsets independently, eliminating the need for a continuous connection with the host system. This is particularly important when operating from a remote site, such as a CFS region, where communication costs are a consideration. Software packages such as RTI's PCLINK (RTI 1985c) and the upcoming "Distributed INGRES" are being investigated to test their ability to turn microcomputers into workstations, permitting distributed processing with the host system. With this arrangement, users can download a selected portion of the database

for browsing, querying, and reporting, or for modification. New data may be entered and edited on the micro before uploading them into the database. A typical workstation that would handle both the distributed database and a CAD/GIS functions is the IBM PC/AT with a 20 MB hard disk, printer, enhanced graphics monitor, 5-pen desk-top plotter, and medium-sized desk-top digitizer. These workstations cost about $15 000 each for the hardware.

System implementation and management

The operation of FIDSINFOBASE will be centred initially at PNFI, with the entire database loaded on a dedicated computer. Before computer access is established at the regional FIDS units, data entry and reporting will be handled out of PNFI. A decentralization of the operation of FIDSINFOBASE will occur eventually out to the regions as FIDS staff become familiar with computer processing and are able to take on these tasks. Terminals connected to regional minicomputers, with communication to the FIDS computer at PNFI, will eventually be used for database access. Augmenting or replacing these terminals with microcomputers for distributed database work will happen when this technique is proven. If desired, the database could also be split up, with regional components residing in their respective minicomputers. These regional databases could be networked for shared access. This option of a distributed database will be investigated, but it is not yet viable because it would place too much load on regional computers. The centralized host database and system management at PNFI, with distributed microcomputer access, is seen as the best solution at present.

Sampling form redesign

Developing the model on the survey sampling form data has helped bring about a better understanding of the true relationships of the information that is collected. The experience permits a redesign of data fields on the form to improve its efficiency in field use, and in data handling. Two levels of redesign can be initiated: modifying data fields without changing the original intent of the survey; and, modifying the information collected specific to the type of survey that is being conducted. The first level is the least complicated, because it involves splitting up data fields, eliminating those that are not used, and adding fields for information that is being collected but not properly recorded. The second level of redesign would involve a systems analysis of the entire FIDS survey to determine recording needs of data not currently provided for on the form. This is required because the current survey form is being used for purposes beyond the intention of its original design. These other data are important but must be kept as separate relations within the database.

Examples of modifications that can be made in a first level of redesign of the survey form are:

A. Inclusion of remarks which explain the collection point as data for the database.

B. Extension of UTM grid designation to the square kilometer level or finer.

C. Splitting of "plot or sample area no." field into two fields: (i) plot type -- whether the plot is temporary or permanent, or data from the Acid

Rain National Early Warning System (ARNEWS) survey; and, (ii) plot number -- which would be unique for the plot type, within the 10 km UTM grid, for that year.

D. Permit entry of the exact figure in the case of quantitative fields. These pertain to basal area, stand height, and number of samples.

E. Modifying the coding system for "covertype" and "dominant species" to a standard which is recognized nationally. The same may apply to the "host tree species" and "land ownership" codes.

F. Modifying "number of units" and "average number" codes to indicate the type of units and items which are being counted.

G. Removing "specimen number" unless it is useful as part of the field survey, or laboratory analysis and creation of a new code. Creation of this new code by extracting from the specimen code the implied data which will identify unique situations such as when samples are not returned from the field, or are sought for in a survey but not found.

H. All survey forms to be preprinted with an unique alphanumeric code which can be used as an identifier for field data, before a "registration number" is later assigned at the laboratory. Note that this assignment could be handled automatically, once the code is entered. Field data could be processed before the lab portion is input.

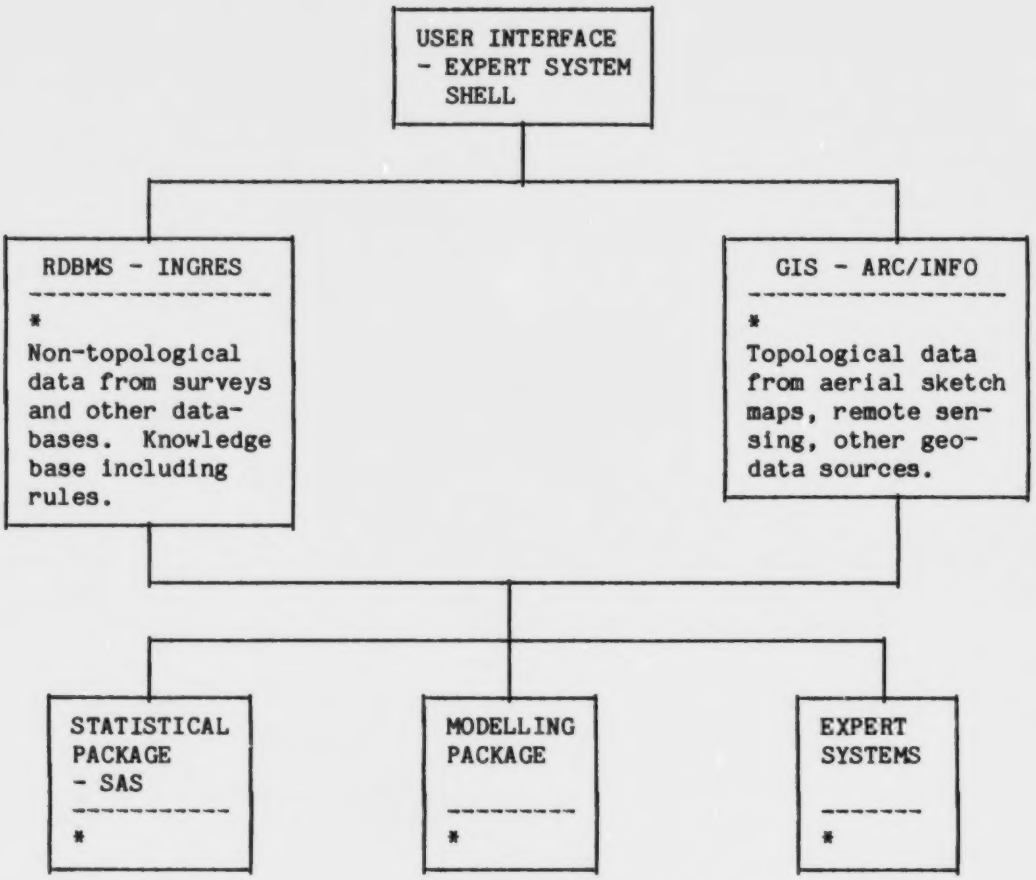
I. All remarks should be viewed as valuable knowledge and entered into the database. It must be recognized that not all observations can be recorded in code, and that a good database must be one that accepts this form of qualitative information.

An integrated information system

Pest databases form the cornerstone from which pest management information systems (PMIS) can be built. A PMIS will process information supporting decisions on pest management issues. The system can be viewed as a "way of doing things" with "tools" to do them. The primary tool in a PMIS will be a computer. PMIS concepts have been studied by other agencies such as the B.C. Ministry of Forests (Sutton et al. 1985) and the USDA Forest Service (Daniel et al. 1983). A goal of the Methods and Applications Group in Forest Pest Management for the USDA Forest Service is to implement a forest pest information system ensuring availability of data for pest management decision-making at all planning levels (USDA Forest Service 1985). FIDSINFOBASE can evolve into a PMIS through the integration of the best in technologies in GIS, RDBMS, Statistical Analysis, and Expert Systems. This concept is shown in Figure 6.

FIDSINFOBASE is not operational with the historical sampling form data, but will be operational with the Acid Rain National Early Warning System (ARNEWS) component in April, 1986. The ARNEWS component will retain all plot data resulting from the ARNEWS survey which is being conducted by FIDS. This component will share relations with the sampling form data, and database operations can be done involving relations from both components. Other data

Figure 6. The Concept of a Forest Pest Management Information System in FIDSINFOBASE.



*to reside on minicomputers, but can also be distributed to microcomputers.

which FIDS has or needs access to can eventually be brought into FIDSINFOBASE, in the same manner as the sampling form and ARNEWS data, such as:

- Parasite studies
- Population surveys
- Depletion/impact surveys and estimates
- Pest collections of insects and diseases, such as those in insectaries and herbariums
- Select forest inventory summary data
- Select weather data
- Protection requirements, activities, and results
- Pest bulletins and reports.

These data may be geographically oriented but are non-topological, and are therefore handled well by the INGRES RDBMS. Topological data would be best handled by a GIS such as ARC/INFO. With FIDSINFOBASE links to a GIS, map data such as pest infestation polygons with physiographic and administrative background data can be retained for graphic display and spatial analysis. The GIS requirements for FIDS have been studied (Williams and Power 1986). These requirements can be used to develop the FIDSINFOBASE GIS components.

In addition to database management and GIS capabilities, a link to a statistical analysis package such as SAS for statistical analysis procedures that the RDBMS and GIS cannot perform, and SASGRAPH for business graphics, would be required. The databases can drive pest condition models to better describe past, current, and future conditions of spread, intensity, and impact. Similar models have been developed for forest fire protection applications (Kourtz 1984).

Expert systems have evolved to a point where they should be seriously considered in any information system. They can be applied for the following purposes:

A. User shell - Any user of the PMIS would enter the system through a shell which asks the user for information on the problem, then directs the user to the components of the system to solve the problem.

B. Knowledge base construction - Much of the data that FIDS has is qualititative and exists on paper or in the minds of both professionals and non-professionals. Expert systems can be used as a mechanism to capture this knowledge and store them in the database in the form of rules.

C. Decision-support - Can assist taxonomic analysis, definition of pest conditions, appraisal of forest impact, decisions on pest and forest management actions, and in survey planning.

CONCLUSION

More efficient computing techniques have made the development of truly useful biological information systems more feasible. FIDS, being an information-oriented agency, can improve its data processing capability and expand the information base necessary for forest pest managers. The development of the first components of FIDSINFOBASE has demonstrated the use of an RDBMS in

quickly producing a working prototype system which is accessible to all users. It is evident that other FIDS data can now be brought into the database. The integration of other data processing systems such as geographic information systems and expert systems will permit FIDSINFOBASE to evolve into a PMIS from which FIDS will get full value from data they collect and the knowledge they have amassed.

ACKNOWLEDGMENTS

The author wishes to thank the following for their kind assistance: FIDS regional and headquarters staff for describing their past, present, and future data types and requirements; DMR and Associates for their part in the feasibility study and prototype development; J. Edwards and L. O'Brien for their development roles; and, L. Magasi, R. Taylor, and D. Leckie for their review of this report.

LITERATURE CITED

Brown, C.E.; Davidson, A.G. 1968. Forest Insect and Disease Survey Data Recording and Retrieval Manual. For. Branch, Dept. Fish. and For., Ottawa, Ont. 71 pp.

Codd, E.F. 1982. Relational Database: A Practical Foundation for Productivity. Comm. ACM. 25(2).

Daniel, T.C.; White, W.B.; Hunter, D.O. 1983. Integrated Pest Impact Assessment System: An interface to a Forest Management Information System. U.S. Dept. of Agriculture, Forest Service, Forest Pest Management Methods and Applications Group, Fort Collins, Co. 22 p.

DMR and Assoc. 1985. Feasibility Study - Pest Specimen Database. Forest Insect and Disease Survey (FIDS) Program - Petawawa National Forestry Institute. DMR and Assoc., Ottawa, Ont. Unpubl. Report. 36 pp.

Hasling, B. 1985. Introduction to Relational Database Technology. The 1985 National Database and Fourth Generation Language Symposium Proceedings. Digital Consulting Associates, Inc., Andover, Mass. Spring H11-H22.

Kent, W. 1983. A Simple Guide to Five Normal Forms in Relational Database Theory. Comm. ACM. 26(2).

Kourtz, P.H. 1984. Decision-Making for Centralized Forest Fire Management. For. Chron. 60(6): 320-327.

Power, J.M.; O'Brien, B.L. 1986. Request for Computer Acquisition at the Petawawa National Forestry Institute for the FIDS Technology Development Project. Can. For. Serv., Petawawa Nat. For. Inst., Chalk River, Ont. Internal Report. 36 pp.

RTI, 1985a. INGRES Database and Application Systems, Vol. 1. Resource Technology Inc., Alameda, Cal. 672 pp.

RTI, 1985b. INGRES Product Overview. Resource Technology Inc., Alameda, Cal. 6 p.

RTI, 1985c. INGRES/PCLINK. Resource Technology Inc., Alameda, Cal. 5 pp.

Russell, B.; Pieronek, J.V. Miller, D.J. Prokrym, T.H.; Rosenbaum, R.R.; Gage, S.H. 1984. Final Report - FY 1983 for Cooperative Agreement No. 12-16-82-001, MSU Dept. Entomol., USDA APHIS-PPQ, Cooperative National Plant Pest Survey and Detection Program. Michigan State University, East Lansing, Mich. 89 pp.

Sandberg, G. 1981. A Primer on Relational Data Base Concepts. IBM Systems Journal. 20(1).

Sutton, R.C.; Wood, P.M.; Shore, T.L.; Cozens, R.D. 1985. Report of the Task Force on a Forest Pest Management Information System for British Columbia. B.C. Forest Service, Forest Protection Branch, Victoria, B.C. 132 pp.

USDA. 1985. FPM/MAG Plan of Work - FY 86. U.S. Dep. of Agriculture, Forest Service, Forest Pest Management Methods and Application Group, Fort Collins, Co. Unpubl. Report. 4 pp.

Williams, D.H.; Power, J.M. 1986. Analysis of Geographic Information Systems Requirements for the Forest Insect and Disease Survey. Can. For. Serv., Petawawa Nat. For. Inst., Chalk River, Ont. (in prep.)

APPENDIX 1. FIDS SURVEY FORM DATA FIELDS

A. Location

i) Collection point - a narrative description of the location of the sampling site.
ii) Locality - a regionally unique administrative unit within which the sampling took place.
iii) UTM grid designation - Universal Transverse Mercator 10 km grid location code.
iv) Plot or Sample Area No. - a regionally unique code which consecutively numbers each plot per UTM. The plot type is inferred in this number. "Plot" is a loose term applying to the locale within which the sampling took place. Certain plot types such as those which are remeasured may have defined boundaries, others do not.
v) Elevation - to nearest 100 ft, or its metric equivalent, above sea level.

B. Stand examined

i) Description - the forest land classification.
ii) History - a description of any past disturbance on the standing timber.
iii) Aspect - one of eight points of the compass which the plot faces.
iv) Maturity - the maturity of the stand.
v) Age structure - a code denoting whether the stand is even, uneven, or two-aged.
vi) Basal area - the class of basal area.
vii) Forest section - Rowe's forest region and section within which plot is located.
viii) Cover or subtype - forest cover type and subtype in terms of tree species and species groups.
ix) Dominant species - a code which identifies the dominant species from the cover type code.
x) Height of dominant trees - class of dominant tree height.

C. Trees sampled

i) Tree species - the host tree species.
ii) No. of trees sampled - actual number of trees sampled.
iii) Status - host tree disturbance.
iv) Tree class - dominance of the host tree.

D. Data on insect(s) or disease(s)

i) Sampling technique - for the qualitative or quantitative sampling technique used in the survey.
ii) Population level or incidence - the presence and concentration of pests in the plot.
iii) Collection source - that part of the host tree or soil from which specimens were taken.
iv) Number of units/average count - the actual number of units of the pest or affected plant material that were sampled, and the average

frequency of an object per sample unit. Examples are: For stem diseases - the number of trees per sample unit and the average number of trees infected per unit; for budworm - the number of trees per sample unit and the number of budworm larvae/m^3; and, for needle rusts - number of branches per unit and the average number of needles infected per unit. These fields are specific to each region and generally requires interpretation by regional staff.

v) Damage hazard index - not currently used.

E. General

i) FIDS no. - a regionally specific code identification of the survey ranger or other cooperator who undertook the survey.
ii) Land ownership - the land ownership class.
iii) Day/Month - day and month in which the survey even took place.
iv) Remarks - uncoded information on the collector, his address, pest symptoms, and other remarks.

F. Form identification

i) Year/region/registration no. - the year and region within which the survey was done is coded. The registration number is a sequential number unique to the region and year, and it is assigned by the regional FIDS unit after a survey.

G. Laboratory analysis - insect collections

i) Specimen no. - a sequential number assigned to each insect species specimen group, beginning with 01, that is included in the collection taken from the survey. Certain numbers are reserved, 00 and 90+, to denote a situation where a pest was specifically sought for in a survey, but not detected, or a situation where a survey on a pest was conducted but no specimens were sent back to the lab for analysis.
ii) Name - the name (normally Latin binomial) of the species as identified by the collector or lab personnel.
iii) Species code - the species as identified following taxonomic procedures.
iv) Stage - the frequency class of insects for each of four possible stages: eggs, larvae, pupae, and adult.
v) Number of insects - the actual number of insects that are contained in the sample for a species from the survey.
vi) Identification - a code which represents the level of expertise of the individual who makes the final identification of the insect.
vii) Disposal - the status of the specimens for the species following identification, such as whether they are discarded or retained in a permanent collection.

H. Laboratory analysis - disease collections

i)-iii) Same as for insect collections
iv) Stage - the stage of development of the disease organism.
v) Symptom - the disease symptoms of the specimen.

vi) Basis of identification - a code which represents the degree of confidence in the identification of the disease.
vii) Identification - same as vi) in insect collections.
viii) Disposal - same as vii) in insect collections.

APPENDIX 2. FIDSINFOBASE RELATIONS (as of early 1986)

Each table or relation is listed with their associated attribute fields, which may be in numeric, alphanumeric, or text types. The row width is the width of the table in character positions. The number of rows indicates the length of the table.

Table: PLOT Row Width: 119 Number of Rows: 0 (will be loaded with 1986 data)

Fields:	Region code	Collection point	UTM zone, east, north
	Plot no.	Locality	Description
	Aspect	Basal area	Elevation
	History	Maturity	Rowe's Section
	Subtype 1	Subtype 2	Dominant species
	Height	Age	Ownership
	Land use		

Table: SURVEY Row Width: 341 Number of Rows: 0 (will be loaded with 1986 data)

Fields:	Region code	Registration no.	UTM zone, east, north
	Survey type	Survey year	Survey date
	Locality	Plot no.	Host type
	Status	Sampling technique	Collection source
	No. of trees	Tree class	Population level
	Sampling units	No. of units	Average count
	Counts per unit	Collector code	Damage index
	Remarks		

Table: SPECIMEN Row Width: 62 Number of Rows: 0 (will be loaded with 1986 data)

Fields:	Survey type	Survey year	Specimen code
	Taxonomic code	No. eggs collected	No. larvae collected
	No. pupae collected	No. adult collected	Total insects
	Dis. devel. stage A	Dis. devel. stage B	Dis. symptom A
	Dis. symptom B	Basis of ID	Identifier
	Disposal	Region code	Registration no.

Table: CFSREGIONS Row width: 166 Number of Rows: 9

Fields:	Region code	Region abbrev.	English name
	French name	Address	City
	Province	Postal code	Phone number

Table: LOCALITY Row width: 76 Number of Rows: 116

Fields:	Region code	Locality	English name
	French name		

Table: ROWESREGION Row width: 97 Number of Rows: 9

Fields: Rowe's region code English name French name
Abbreviation

Table: ROWESECTION Row Width: 99 Number of Rows: 43

Fields: Rowe's region code Section code English name
French name Abbreviation

Table: OWNERSHIPS Row Width: 73 Number of Rows: 9

Fields: Ownership code English name French name

Table: ASPECTS Row Width: 25 Number of Rows: 10

Fields: Aspect code English name French name
English abbrev. French abbrev.

Table: HISTORIES Row Width: 33 Number of Rows: 12

Fields: History code English name French name

Table: AGESTRUCTURE Row Width: 41 Number of Rows: 4

Fields: Age code English name French name

Table: DESCRIPTIONS Row Width: 49 Number of Rows: 11

Fields: Description code English name French name

Table: MATURITIES Row Width: 61 Number of Rows: 9

Fields: Maturity code English name French name

Table: STATUS Row Width: 61 Number of Rows: 8

Fields: Status code English name French name

Table: LANDUSE Row Width: 73 Number of Rows: 10

Fields: Land use code English name French name

Table: HOSTTABLE Row Width: 145 Number of Rows: 53

Fields: Host type Host group Latin name
Eng. common name Fr. common name Abbreviation

Table: SUBTYPE Row Width: 45 Number of Rows: 43

Fields: Subtype code Name Abbreviation

Table: TREECLASS Row Width: 61 Number of Rows: 10

Fields: Tree class code English name French name

Table: SAMPTECH Row Width: 61 Number of Rows: 13

Fields: Sampling tech. code English name French name

Table: COLLECTSRC Row Width: 25 Number of Rows: 13

Fields: Collection source code English name French name

Table: POPLEVELS Row Width: 33 Number of Rows: 7

Fields: Population level code English name French name

Table: COLLECTORS Row Width: 196 Number of Rows: 0 (to be entered)

Fields: Region code Collector code Establishment
Name Address City
Province Postal code Phone number

Table: CAUSALAGENTS Row Width: 84 Number of Rows: 22, 140

Fields: Taxonomic code Class/genus Species
Change type Catalogue type

Table: SYNONYM Row Width: 17 Number of Rows: 6193

Fields: Taxonomic code Synonym Catalogue type
Synonym flag 1 Synonym flag 2

Table: DISDEVATAB Row Width: 73 Number of Rows: 11

Fields: Dis. Devel. A English name French name

Table: DISDEVATAB Row Width: 73 Number of Rows: 10

Fields: Dis. Devel. B English name French name

Table: BASISID Row Width: 73 Number of Rows: 7

Fields: Basis ID code English name French name

Table: IDENTIFIER Row Width: 73 Number of Rows: 7

Fields: Identifier code English name French name

Table: DISPOSAL Row Width: 73 Number of Rows: 12

Fields: Disposal code English name French name

Table: AUTHORITY Row Width: 28 Number of Rows: 99

Fields: User ID Function

Table: FUNCTION Row Width: 80 Number of Rows: 51

Fields: Program Function Description

Tables: REG*HISTORY where * is a code for each of the six CFS regions.

Region	Row Width	Number of Rows
Pacific	157	>439,000
Northern	157	>108,000
Great Lakes	157	>273,000
Laurentian	157	>108,000
Maritimes	157	>223,000
Newfoundland	157	>56,000

Fields:	Survey type	Region code	Survey year
	Registration no.	Locality	UTM zone, east, north
	Plot no.	Elevation	Description
	History	Aspect	Maturity
	Basal area	Basal code	Section code
	Subtype 1	Subtype 2	Dominant species
	Height code	Height	Age code
	Ownership code	Land use	Host type
	No. of trees	Status	Tree class
	Sampling technique	Population level	Collection source code
	Sample unit	No. of units	Count per unit
	Average count	Damage index	Collector code
	Survey day	Survey month	Survey year
	Survey date	Specimen code	Interpreter
	Taxonomic code	Insect stage	No. of insects
	No. eggs collected	No. larvae collected	No. pupae collected
	No. adults collected	Total insects	Dis. devel. A
	Dis. devel. B	Dis. symptom A	Dis. symptom B
	Disposal	Basis ID	Identification code